MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 4689–4691

NEWS AND VIEWS

PERSPECTIVE

Making sense of genetic estimates of effective population size

ROBIN S. WAPLES

NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

The last decade has seen an explosion of interest in use of genetic markers to estimate effective population size, N_e. Effective population size is important both theoretically (N_e is a key parameter in almost every aspect of evolutionary biology) and for practical application (N_e determines rates of genetic drift and loss of genetic variability and modulates the effectiveness of selection, so it is crucial to consider in conservation). As documented by Palstra & Fraser (2012), most of the recent growth in N_e estimation can be attributed to development or refinement of methods that can use a single sample of individuals (the older temporal method requires at least two samples separated in time). As with other population genetic methods, performance of new N_e estimators is typically evaluated with simulated data for a few scenarios selected by the author(s). Inevitably, these initial evaluations fail to fully consider the consequences of violating simplifying assumptions, such as discrete generations, closed populations of constant size and selective neutrality. Subsequently, many researchers studying natural or captive populations have reported estimates of N_e for multiple methods; often these estimates are congruent, but that is not always the case. Because true N_e is rarely known in these empirical studies, it is difficult to make sense of the results when estimates differ substantially among methods. What is needed is a rigorous, comparative analysis under realistic scenarios for which true N_e is known. Recently, Gilbert & Whitlock (2015) did just that for both single-sample and temporal methods under a wide range of migration schemes. In the current issue of Molecular Ecology, Wang (2016) uses simulations to evaluate performance of four single-sample N_e estimators. In addition to assessing effects of true N_e, sample size, and number of loci, Wang also evaluated performance under changing abundance, physical linkage and genotyping errors, as well as for some alternative life histories (high rates of selfing; haplodiploids). Wang showed that the sibship frequency (SF) and linkage disequilibrium (LD) methods perform dramatically better than the heterozygote excess and molecular coancestry methods under most scenarios (see Fig. 1, modified from figure 2 in Wang 2016), and he also concluded that SF is generally more versatile than LD. This article represents a truly Herculean effort, and results should be of considerable value to researchers interested in applying these methods to real-world situations.

Correspondence: Robin S. Waples, Fax: +1 206 860-3335; E-mail: robin.waples@noaa.gov

Keywords: bias, computer simulations, linkage disequilibrium, precision, siblings

Received 1 August 2016; revision received 12 August 2016; accepted 16 August 2016

Still, Wang's paper is not likely to be the last word on single-sample estimators, for several reasons. (i) Although the range of scenarios considered is truly impressive, it is not comprehensive. For example, most real populations have overlapping generations and are connected to other populations by migration, but neither was modelled in Wang (2016). (ii) Nonrandom samples (with an excess of relatives) appear to be very common in studies of natural populations, but the consequences of this for estimating N_e have not been systematically evaluated in any study. It seems likely the SF method would be particularly sensitive to this issue. (iii) The prior for N_e introduced by Wang in this article appears to be potentially very useful but needs fuller evaluation, as do two empirical correction factors he derived that are used in calculating the prior. Except in the set of simulations evaluating effects of different priors, it seems that the true N_e was used as the prior for the SF method. This would improve performance with simulated data, but using true N_e as a prior is not an option for a researcher who wants to find out what the true N_e is. (iv) One promising single-sample estimator using approximate Bayesian computation (ONeSAMP; Tallmon *et al.* 2008) was not included in Wang's paper. Use of ONeSAMP has been limited by an online-only implementation, and it was temporarily unavailable in early 2016. Downloadable and online versions of ONeSAMP 2.0, which can accept SNP or mSAT data, are now available at http://www.cs.colostate.edu/~onesamp/index.php.

It is inevitable that in the future an increasing fraction of genetic estimates of N_e will be based on genomics-scale data sets with many thousands of loci, and there are some interesting tradeoffs between the SF and LD methods in this respect that were not fully evaluated by Wang. Figure 1 shows that when only a few loci are available, the SF method performs relatively well, provided sample sizes are adequate and true N_e is not very large. However, performance of the SF method actually declined slightly when more than 10 'microsat' loci were used. Presumably this reflects influence of prior, as well as the fact that once sibships are correctly identified, more markers do not

Published 2016. This article is a U.S. Government work and is in the public domain in the USA

FWS_LIT_032123

improve performance of the SF method. In contrast, the root-mean-square error of $1/\hat{N}_e$ (RMSE) for the LD method declined steadily with increasing loci (Fig. 1), suggesting that performance would exceed that of the SF method if more loci were used. However, although it is easy to model lots of 'unlinked' markers in a computer, real organisms have a limited number of chromosomes, so physical linkage is inevitable, and this linkage downwardly biases $\hat{N}_e$ for the LD method. As it happens, another recent paper (Waples *et al.* 2016) showed that this downward bias in $\hat{N}_e$ in the LD method is related to recombination rate and can be estimated from the number of chromosomes (or the total genome length, as modelled by Wang). This suggests that a simple bias correction for linkage could enable the LD method to take advantage of increased precision afforded by genomics-scale data sets.

The papers by Gilbert & Whitlock (2015) and Wang (2016) share another feature besides comparative evaluation of N_e-estimation methods: they both were conducted by authors of likelihood-based software that is very computationally demanding (estimates for a single sample can take hours, days or even weeks). Although other researchers routinely use these programs for analysis of empirical data, it would be a major undertaking for any third party to rigorously evaluate performance using simulated data for hundreds or thousands of replicates for each scenario of interest (by my count, Wang evaluated two metrics – harmonic mean $\hat{N}_e$ and RMSE – for over 55 different parameter combinations for each of four estimators). As a consequence, most such evaluations are carried out by the authors themselves, who are much better positioned to develop streamlined pipelines to conduct the daunting matrix of analyses. I am not suggesting that either of the above papers is anything but objective, and (disclaimer!) I have been involved with a number of published evaluations of temporal or LD methods that I helped develop. Furthermore, this issue is not restricted to N_e estimators; it also applies to many other software packages for population genetic data analysis. However, anything that facilitates independent, third-party evaluation of such programs will benefit the scientific community as a whole. I do not have any simple solutions to offer, but some guidelines could help. Programs should accept input files in standard formats. Slick graphical interfaces are fine, provided an option also exists for batch-processing large numbers of replicate samples. Output files should be in machine-readable formats. Simulation code should be made available on a public site. It is encouraging to see that Wang (2016) has posted on Dryad code and instructions for conducting the types of simulations and analyses reported in his paper.

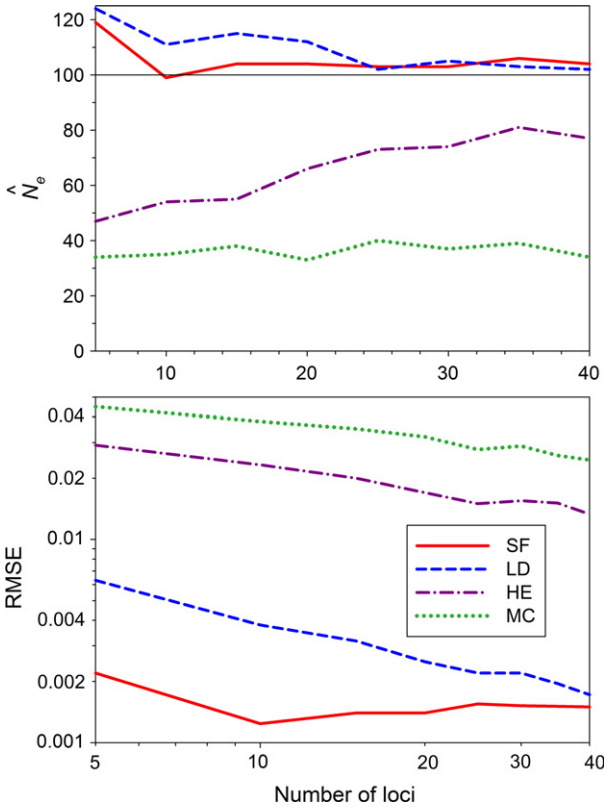

Fig. 1 Performance of four single-sample estimators with simulated data for an isolated population with true $N_e = 100$ (thin black line in top graph), as a function of the number of unlinked 'microsat' loci (each with 10 alleles). Sample size was 50 individuals. Note the log scale for both axes in the bottom graph. Methods: SF = sibship frequency (Wang 2009); LD = linkage disequilibrium (Waples & Do 2008); HE = heterozygote excess (Pudovkin *et al.* 1996); MC = molecular coancestry (Nomura 2008). Modified from figure 2 in Wang (2016).

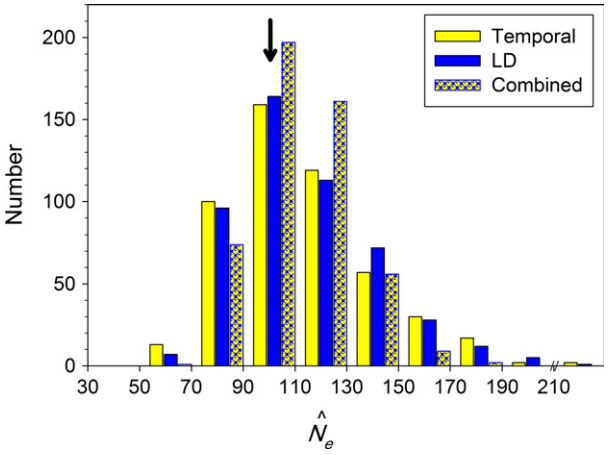

Fig. 2 Distribution of 500 estimates of effective size for the temporal (Waples 1989) and LD (Waples & Do 2008) methods, based on simulated data with constant $N_e = 100$ (arrow) and samples of 50 individuals genotyped for 100 unlinked and diallelic (SNP) loci. The temporal samples were taken six generations apart, and only the second sample was used for the LD analyses. For each replicate, a 'combined' estimate was also computed as the harmonic mean of $\hat{N}_e$ for the temporal and LD methods. Data were simulated as described by Waples (in press).

Published 2016. This article is a U.S. Government work and is in the public domain in the USA

FWS_LIT_032124

Finally, I would like to highlight a potential opportunity to improve our ability to estimate effective size in natural populations by combining information from more than one estimator. If two methods with approximately comparable performance provide an estimate of a parameter of interest, the variance of a combined estimate will be smaller than for either estimate alone, providing the estimates are not strongly positively correlated. This principle can be illustrated by combining estimates of N_e from simulated data using the temporal and LD methods, which provide essentially independent information (Fig. 2). Under the scenario modelled, the two estimators had nearly the same precision (CV of $\hat{N}_e$ = 0.26 and 0.25 for the temporal and LD methods, respectively), and the 500 replicate $\hat{N}_e$ values for the two estimators were uncorrelated ($r = -0.001$). Combining the estimates (in this case by taking an unweighted harmonic mean $\hat{N}_e$) reduced the CV to 0.17 and visibly shrunk the distribution of $\hat{N}_e$ (Fig. 2). The LD and SF methods also use largely independent information, so the estimates they produce might be uncorrelated as well, but that is only a conjecture that requires empirical evaluation. Questions to explore include: (i) What is the optimal way to combine results from different estimators? Something like a weighted harmonic mean of $\hat{N}_e$ might be best, with the weights being an inverse function of RMSE. (ii) How could one compute confidence intervals for a combined estimate?

Acknowledgements

Jinliang Wang kindly provided the data for Fig. 1. I thank Peter Beerli, David Tallmon and Ryan Waples for useful comments and discussion.

References

Gilbert KJ, Whitlock MC (2015) Evaluating methods for estimating local effective population size with and without migration. *Evolution*, **69**, 2154–2166.

Nomura T (2008) Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications*, **1**, 462–474.

Palstra FP, Fraser DJ (2012) Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution*, **2**, 2357–2365.

Pudovkin AI, Zaykin DV, Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*, **144**, 383–387.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) OneSamp: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, **8**, 299–301.

Wang J (2009) A new method for estimating effective population size from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148–2164.

Wang J (2016) A comparison of single-sample estimators of effective population sizes from genetic marker data. *Molecular Ecology*, **25**, 4692–4711.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.

Waples RS (In press) Tiny Ne/N ratios in marine species: are they real? *Journal of Fish Biology*, **89**; doi 10.1111/jfb.13143.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753–756.

Waples RK, Larson WA, Waples RS (2016) Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*,doi:10.1038/hdy.2016.60. [Epub ahead of print]

R.S.W. conceived the study, conducted the analyses, and wrote the paper.

---

doi: 10.1111/mec.13814

Published 2016. This article is a U.S. Government work and is in the public domain in the USA

FWS_LIT_032125